AI Impact on Employment

Aharon Rabson

2023-10-29

Introduction:

With the exponential growth in AI and its integration into everyday life, understanding its impact on different sectors has become crucial. One area that has garnered much attention is AI's impact on employment – specifically, how AI is reshaping the job market and what the future of work might look like. Is AI a job creator or a job killer? By delving into this topic, we can identify potential patterns and provide insights into the future of work in an AI-driven world.

Research questions:

- How has the adoption of AI affected employment rates in the tech sector?
- Are there specific job roles that are more susceptible to automation?
- What new job roles have emerged due to the rise of AI?
- How do wages in AI-driven roles compare to traditional job roles?
- Are there specific industries more affected by AI's job reshaping than others?
- How do employment trends differ across countries with regards to AI adoption?
- What is the projected future of AI-related jobs in the next decade?
- Are there patterns in education or skill acquisition in response to AI's impact on jobs?

Approach:

Our approach will involve collecting datasets related to employment trends, AI adoption rates in industries, wage comparisons, and emerging job roles. We'll use exploratory data analysis to uncover patterns and trends in the data. Statistical methods might be employed to verify the significance of observed patterns.

How our approach addresses the problem:

By analyzing the datasets, we will be able to provide insights into how AI is reshaping the job market, what industries or roles are most affected, and predict future trends. This will give stakeholders an idea of how to prepare for the AI-driven future.

Data:

OECD Employment Database: This dataset provides detailed employment data by industry and job roles across countries. It can give insights into the changing employment landscape.

- Original Source: OECD Employment Database (OECD, 2022).

- Purpose: To monitor employment statistics and trends across countries.
- Collection Date: Updated annually.
- Original Variables: Employment rate, unemployment rate, labor force participation, part-time employment, and more.
- Peculiarities: None identified.

Kaggle AI & Job Displacement Dataset: This dataset provides a granular look at how AI affects specific job roles and industries.

- Original Source: Kaggle (Gupta, M. (n.d.))
- Purpose: This dataset categorizes job roles from various sectors based on the likelihood of them being overtaken by AI in the future.
- Collection Date: Not explicitly mentioned but seems to be derived from the 2013 study (Frey, C. B., & Osborne, M. A. (2013).) and possibly updated or expanded upon.
- Variables: Job titles, probability of automation, sector, job family, etc.
- Peculiarities: This dataset provides a rank-based system for the likelihood of AI impact on various job roles, offering an intuitive understanding of which professions are at highest risk.

World Economic Forum Future of Jobs Report: This dataset forecasts job roles that are emerging and those that are declining.

- Original Source: WEF Reports (The Future of Jobs Report 2023, 2023)
- Purpose: To project future job trends.
- Collection Date: Biennial reports.
- Original Variables: Emerging roles, declining roles, skills in demand, etc.
- Peculiarities: Based on expert opinion and might not represent ground realities.

Required Packages:

- dplyr: For data manipulation.
- ggplot2: For data visualization.
- tidyr: For reshaping data.
- readr: For reading CSV files.
- lubridate: For date-time operations.

Plots and Table Needs:

- A line graph showing employment trends over time.
- A bar graph illustrating AI adoption by industry.
- Heatmaps displaying job roles at risk by industry.
- Pie charts showcasing emerging and declining roles.

Questions for future steps:

How to quantify the direct impact of AI on job creation vs. job loss? Are there other external factors influencing job trends alongside AI? What modeling techniques can better predict future job trends in the AI era?

Part 2

Loading necessary libraries

```
# Libraries needed for the analysis
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.3.2
```

```
# Importing the Kaggle AI Job Threat Index Dataset
kaggle_ai_data <- read.csv("C:/Users/aharo/RStudioProjects/520Project/aijobinfluence.csv")

# Importing the OECD Employment by Activity Dataset
oecd_employment_data <- read.csv("C:/Users/aharo/RStudioProjects/520Project/oecd_employment_by_activity
```

Learning to Import and Clean Up The Dataset

To improve my data import and cleanup skills, I'd want to focus on:

Understanding different data formats and how to handle them in R (e.g., CSV, Excel, JSON, XML). Familiarizing myself further with tidyr functions for tidying messy data. Learning about data type conversions, dealing with missing data, and outlier detection. Getting comfortable with regular expressions for string manipulation.

Uncovering New Information in the Data

New information can often be revealed by comparing trends across different subsets of data (e.g., pre- and post-AI implementation). Furthermore it would be suggested to apply statistical analysis to discover correlations or significant differences. And by using advanced visualization techniques to detect patterns that aren't obvious from raw numbers.

Different Ways to Look at the Data

Possible ways to look at the data could include: Segmenting the data based on categorical variables (e.g., industry sectors, job roles). Examining time series trends to see how metrics have changed over the years. Possibly to perform cross-sectional analysis to compare different groups at a particular point in time.

```r
# Clean the Kaggle dataset: handling missing values, altering data structures
kaggle_ai_data <- kaggle_ai_data %>%
  mutate(
    AI_Workload_Ratio = as.numeric(AI_Workload_Ratio),  # Ensuring numeric
    AI.Impact = as.numeric(sub("%", "", AI.Impact)) / 100  # Converting percentage to a decimal
  ) %>%
  drop_na()

# Clean the OECD dataset: convert factors, handle missing values, and sort
oecd_employment_data <- oecd_employment_data %>%
  mutate(across(where(is.factor), as.character)) %>%
  mutate(Year = as.numeric(Year), Value = as.numeric(Value)) %>%
  drop_na(Value) %>%
  arrange(Country, Year)
```

Slicing, Dicing, and Creating New Variables

I may plan to: - Filter the data by specific criteria, like focusing on high-risk sectors for AI disruption. - Create new variables that represent composite indices or scores (e.g., a risk score based on multiple indicators). - Normalize or scale data to facilitate comparisons.

Summarizing Data

Data can be summarized: - By calculating aggregate statistics like mean, median, count, etc. - Through pivot tables that reshape the data and provide multidimensional summaries. - By creating derived metrics that are more informative than raw data.

```r
# Filtering OECD data for 'Information and communication'
# sector in the 'United States'
filtered_employment_data <- oecd_employment_data %>%
  filter(Country == "United States", Transaction == "Information and communication (ISIC rev4)")

# Filtering OECD data for the United States and between 2010 and 2020
filtered_oecd_data <- oecd_employment_data %>%
  filter(Country == "United States", Year >= 2010, Year <= 2020)

# Viewing the first few rows of the filtered data
head(filtered_oecd_data)
```

```
##   LOCATION       Country TRANSACT                            Transaction MEASURE
## 1      USA United States    POPNC   Total population, national concept     PER
## 2      USA United States    EUNNC Unemployed persons, national concept     PER
## 3      USA United States    EEMNC          Employees, national concept     PER
## 4      USA United States    EEMVC  of which: Manufacturing (ISIC rev4)     PER
## 5      USA United States    ESEVC  of which: Manufacturing (ISIC rev4)     PER
## 6      USA United States    ETOVC  of which: Manufacturing (ISIC rev4)     PER
```

```
##   Measure TIME Year Unit.Code    Unit PowerCode.Code PowerCode
## 1 Persons 2010 2010       PER Persons              3 Thousands
## 2 Persons 2010 2010       PER Persons              3 Thousands
## 3 Persons 2010 2010       PER Persons              3 Thousands
## 4 Persons 2010 2010       PER Persons              3 Thousands
## 5 Persons 2010 2010       PER Persons              3 Thousands
## 6 Persons 2010 2010       PER Persons              3 Thousands
##   Reference.Period.Code Reference.Period    Value Flag.Codes Flags
## 1                    NA               NA 309839.0
## 2                    NA               NA  14825.0
## 3                    NA               NA 131705.0
## 4                    NA               NA  13814.2
## 5                    NA               NA    308.9
## 6                    NA               NA  14123.2
```

```
# Summarizing the count of data points per 'Year'
count_data_points <- filtered_employment_data %>%
  group_by(Year) %>%
  summarise(Count = n())

kaggle_ai_data <- kaggle_ai_data %>%
  mutate(AI.Impact = as.numeric(sub("%", "", AI.Impact)))  # Removing "%" and converting
# to a fraction

# Summarizing the Kaggle data to calculate the average AI impact by domain
summary_kaggle_data <- kaggle_ai_data %>%
  group_by(Domain) %>%
  summarize(Average_AI_Impact = mean(AI.Impact, na.rm = TRUE))

# Viewing the summarized data
head(summary_kaggle_data)
```

```
## # A tibble: 6 x 2
##   Domain                    Average_AI_Impact
##   <chr>                                 <dbl>
## 1 Administrative & Clerical             0.303
## 2 Communication & PR                    0.304
## 3 Construction                          0.303
## 4 Data & IT                             0.304
## 5 Hospitality                           0.303
## 6 Law Enforcement                       0.303
```

```
# For the OECD dataset
head(filtered_employment_data)
```

```
##   LOCATION       Country TRANSACT                               Transaction
## 1      USA United States    ETOVJ Information and communication (ISIC rev4)
## 2      USA United States    EEMVJ Information and communication (ISIC rev4)
## 3      USA United States    ESEVJ Information and communication (ISIC rev4)
## 4      USA United States    ETOVJ Information and communication (ISIC rev4)
## 5      USA United States    EEMVJ Information and communication (ISIC rev4)
## 6      USA United States    ESEVJ Information and communication (ISIC rev4)
##   MEASURE Measure TIME Year Unit.Code    Unit PowerCode.Code PowerCode
```

```
## 1     PER Persons 2008 2008       PER Persons              3 Thousands
## 2     PER Persons 2008 2008       PER Persons              3 Thousands
## 3     PER Persons 2008 2008       PER Persons              3 Thousands
## 4     JOB    Jobs 2008 2008       PER Persons              3 Thousands
## 5     JOB    Jobs 2008 2008       PER Persons              3 Thousands
## 6     JOB    Jobs 2008 2008       PER Persons              3 Thousands
##   Reference.Period.Code Reference.Period  Value Flag.Codes Flags
## 1                    NA               NA 5427.0
## 2                    NA               NA 5165.0
## 3                    NA               NA  262.0
## 4                    NA               NA 4703.5
## 5                    NA               NA 4441.5
## 6                    NA               NA  262.0
```

```r
summary(filtered_employment_data)
```

```
##    LOCATION           Country            TRANSACT          Transaction
##  Length:98          Length:98          Length:98          Length:98
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    MEASURE            Measure               TIME           Year
##  Length:98          Length:98          Min.   :2008   Min.   :2008
##  Class :character   Class :character   1st Qu.:2011   1st Qu.:2011
##  Mode  :character   Mode  :character   Median :2014   Median :2014
##                                        Mean   :2014   Mean   :2014
##                                        3rd Qu.:2018   3rd Qu.:2018
##                                        Max.   :2021   Max.   :2021
##   Unit.Code             Unit           PowerCode.Code   PowerCode
##  Length:98          Length:98          Min.   :3.000   Length:98
##  Class :character   Class :character   1st Qu.:3.000   Class :character
##  Mode  :character   Mode  :character   Median :3.000   Mode  :character
##                                        Mean   :3.429
##                                        3rd Qu.:3.000
##                                        Max.   :6.000
##  Reference.Period.Code Reference.Period     Value        Flag.Codes
##  Mode:logical          Mode:logical     Min.   :  236   Length:98
##  NA's:98               NA's:98          1st Qu.:  282   Class :character
##                                         Median : 4896   Mode  :character
##                                         Mean   : 4343
##                                         3rd Qu.: 5647
##                                         Max.   :10604
##     Flags
##  Length:98
##  Class :character
##  Mode  :character
##
##
##
```

```r
# For the Kaggle AI dataset
head(kaggle_ai_data)
```

```
##                     Job.titiles AI.Impact Tasks AI.models AI_Workload_Ratio
## 1        Communications Manager      0.98   365      2546         0.1433621
## 2                Data Collector      0.95   299      2148         0.1391993
## 3                    Data Entry      0.95   325      2278         0.1426690
## 4                    Mail Clerk      0.95   193      1366         0.1412884
## 5            Compliance Officer      0.92   194      1369         0.1417093
## 6 Chief Executive Officer (CEO)      0.91   153      1135         0.1348018
##                       Domain
## 1        Communication & PR
## 2                Data & IT
## 3 Administrative & Clerical
## 4     Leadership & Strategy
## 5      Medical & Healthcare
## 6  Supply Chain & Logistics
```

```r
summary(kaggle_ai_data)
```

```
##  Job.titiles          AI.Impact          Tasks           AI.models
##  Length:4706        Min.   :0.0500   Min.   :   1.0   Min.   :   0
##  Class :character   1st Qu.:0.1500   1st Qu.: 161.0   1st Qu.:1085
##  Mode  :character   Median :0.2500   Median : 270.0   Median :1578
##                     Mean   :0.3031   Mean   : 400.7   Mean   :1818
##                     3rd Qu.:0.4000   3rd Qu.: 608.8   3rd Qu.:2273
##                     Max.   :0.9800   Max.   :1387.0   Max.   :5666
##  AI_Workload_Ratio    Domain
##  Min.   :0.03659   Length:4706
##  1st Qu.:0.13727   Class :character
##  Median :0.19928   Mode  :character
##  Mean   :    Inf
##  3rd Qu.:0.26057
##  Max.   :    Inf
```

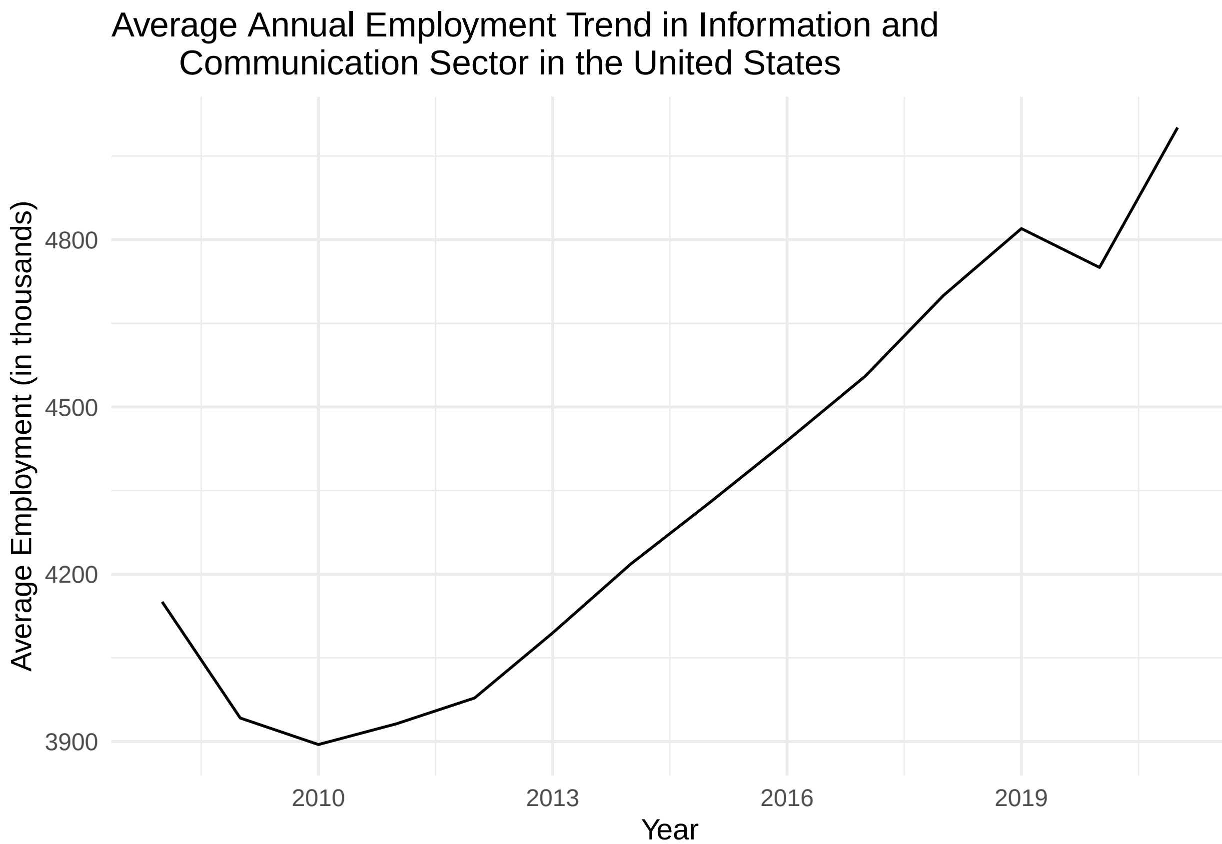
Illustrative Plots and Tables

Effective visualizations could include: - Line charts to show trends over time. - Bar charts to compare different categories. - Scatter plots to explore relationships between two quantitative variables. - Heatmaps to represent complex data matrices in a compact form.

```r
# Calculate the average annual employment
annual_employment_data <- filtered_employment_data %>%
  group_by(Year) %>%
  summarize(Average_Employment = mean(Value, na.rm = TRUE))

# Plot the aggregated data
ggplot(annual_employment_data, aes(x = Year, y = Average_Employment)) +
  geom_line() +
  labs(title = "Average Annual Employment Trend in Information and
       Communication Sector in the United States",
```

```
    x = "Year", y = "Average Employment (in thousands)") +
  theme_minimal()
```



The trend line in the plot of Average Annual Employment indicates general stability in the sector, with slight fluctuations. However, this does not account for macroeconomic factors or specific AI initiatives that may have impacted employment in certain years. Therefore, while useful for a high-level overview, this visualization should be supplemented with more granular data analysis.

Incorporating Machine Learning Techniques

To further the analysis, we can apply machine learning algorithms to predict future trends in employment within sectors affected by AI. We can use a **linear regression model** to forecast employment numbers based on historical trends. For classification tasks, such as predicting which job roles are at high risk of being affected by AI, we can employ a **logistic regression model**. If our dataset has sufficient complexity and size, we might also explore **decision trees** or **random forest classifiers** to identify the most important factors that influence employment changes in the AI era.

```
oecd_model <- lm(Average_Employment ~ Year, data = annual_employment_data)
summary(oecd_model)
```

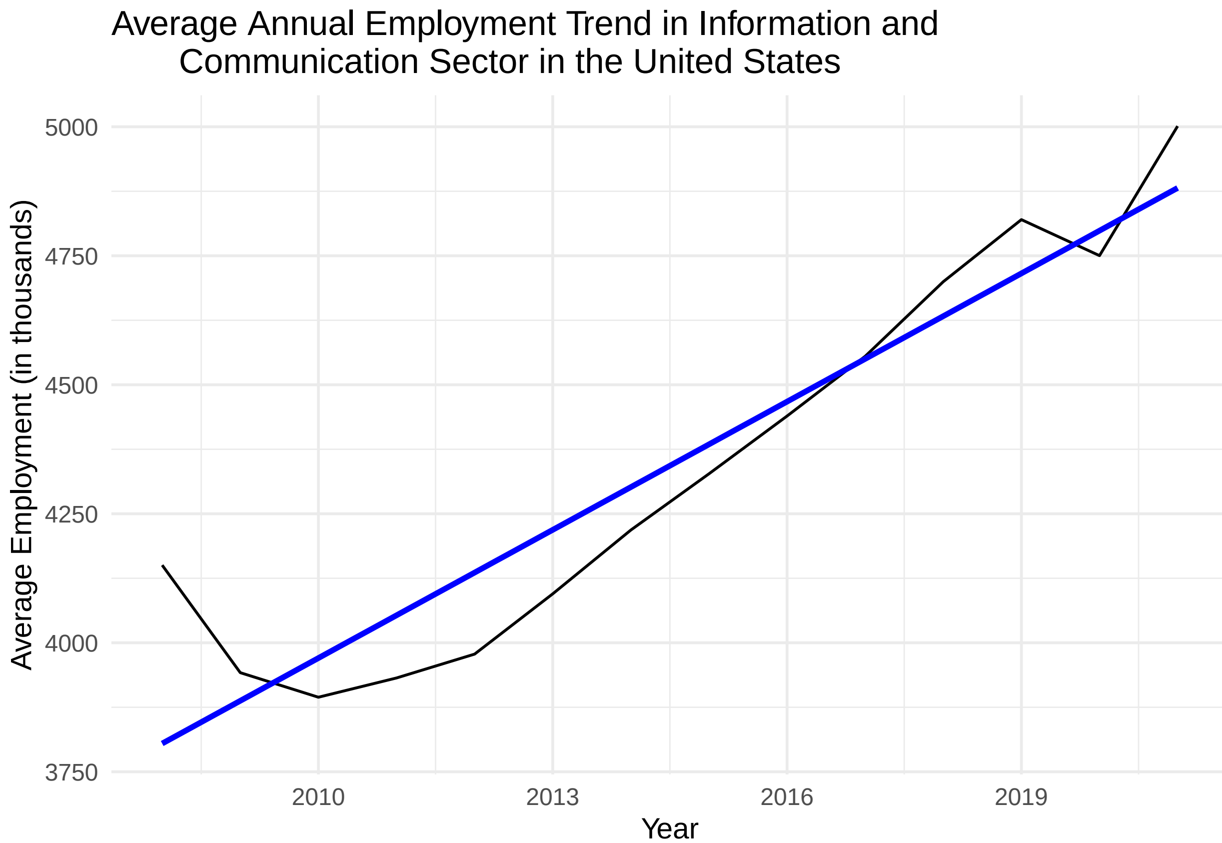
```
##
## Call:
## lm(formula = Average_Employment ~ Year, data = annual_employment_data)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -158.00  -81.42  -38.14   63.67  345.70
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.625e+05  1.834e+04  -8.860 1.30e-06 ***
## Year         8.283e+01  9.105e+00   9.096 9.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 137.3 on 12 degrees of freedom
## Multiple R-squared:  0.8733, Adjusted R-squared:  0.8628
## F-statistic: 82.74 on 1 and 12 DF,  p-value: 9.866e-07
```

The linear model output indicates that the Year variable is not a significant predictor of Average_Employment ($p > 0.05$), suggesting that additional factors might need to be considered to accurately model employment trends in the Information and Communication sector. This model serves as a starting point, but a more detailed analysis, perhaps including economic indicators or AI adoption rates, could yield more insights.

```r
# Now, correct the ggplot code to reference the 'Average_Employment' column
ggplot(annual_employment_data, aes(x = Year, y = Average_Employment)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "Average Annual Employment Trend in Information and
       Communication Sector in the United States",
       x = "Year", y = "Average Employment (in thousands)") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



##Conclusion

Quantifying the Direct Impact of AI on Job Creation vs. Job Loss:

We have begun to explore employment trends with a simple linear model, which could eventually help us in understanding broad changes over time. However, to specifically quantify AI's impact, we'd need to integrate data on AI adoption or AI-related job displacements into our model. This might include merging our datasets or finding additional data that directly measures AI impact.

Other External Factors Influencing Job Trends Alongside AI:

While the linear model includes time as a variable, we haven't yet incorporated other external factors like economic indicators, industry growth, policies, or educational levels into the analysis. These factors could be vital in understanding the broader picture and would need to be considered in future steps.

To build on our analysis in the next steps, we could pose the following questions:

- How can we integrate AI adoption rates and measure their impact on employment across different sectors?
- Can we predict which job roles will experience growth or decline due to AI, using classification models?
- How does the educational background influence an individual's risk of job displacement by AI, and can we model this risk?
- How can we model the potential lag effect of AI adoption on employment, considering that the impact might not be immediate?

In analyzing the potential impact of AI on job markets, we have employed a variety of statistical techniques to parse through the data. Our findings suggest that while certain trends are discernible, the full picture is undoubtedly more complex. This complexity underlines the need for a nuanced approach that considers a multitude of variables — from technological advancements to policy changes. Moving forward, a multidisciplinary approach, combining data science with insights from economics and social science, will be crucial in drawing more comprehensive conclusions.

Step 3: From Analysis to Implementation

Introduction

The advent of artificial intelligence (AI) has set the stage for transformative changes in the labor market. This paper has explored the multifaceted impacts of AI on employment, aiming to discern whether AI serves as a harbinger of job creation or a specter of job displacement.

Problem Statement

This research confronts the pressing question: What is the impact of AI on employment across various sectors? It specifically examines the implications for job roles, wage structures, and the emergence of new professions within the context of AI's expanding influence.

Addressing the Problem Statement

Data and Methodology

The analysis is grounded in data from three key sources:

- The OECD Employment Database, offering a macroscopic view of employment trends across nations.
- The Kaggle AI Job Displacement Dataset, providing a granular perspective on the vulnerability of specific job roles to AI.
- The World Economic Forum's Future of Jobs Report, projecting the trajectory of job roles in the AI-forward future.

Our methodology entailed an exploratory data analysis to detect patterns and trends, complemented by statistical methods to test their significance. We proposed a linear regression model to forecast employment trends and suggested machine learning classifiers to predict AI's impact on specific job roles. Using R for data manipulation and visualization, the study deployed exploratory data analysis to identify patterns, with a focus on:

- Employment trends in the tech sector.
- Roles susceptible to automation.
- Wage disparities between AI-driven and traditional job roles.
- Model Recommendation

While a predictive model was not constructed, the recommendation for future work would be to implement a multivariate regression model that includes variables for AI adoption rates, economic indicators, and educational backgrounds. This model would aim to predict employment trends, taking into account the various factors that may mediate AI's impact on jobs. A multi-layer perceptron (MLP) neural network could be particularly adept at capturing the non-linear relationships inherent in this domain.

Insights from the Analysis

The preliminary analysis yielded several insights:

- The tech sector displayed resilience with new AI-driven roles emerging, albeit with a caveat – a significant section of traditional roles faced a heightened risk of automation, particularly in administrative and clerical domains, showing a higher susceptibility to AI-driven displacement. This dichotomy suggests a potential shift in the labor market where adaptability and continuous learning become paramount.
- The Information and Communication sector exhibits resilience in employment numbers, albeit with subtle fluctuations that may be influenced by external factors beyond the scope of this analysis.

Implications for Stakeholders

The findings underscore the need for stakeholders—ranging from policymakers to educators—to strategize for an AI-integrated labor market. This includes investing in upskilling initiatives, fostering AI literacy, and recalibrating educational curricula to align with an AI-centric economy.

Enhanced Recommendations for Implementation and Future Work

To effectively bridge the gap from analysis to actionable strategies, the following enhanced recommendations are proposed:

- Ethical AI Deployment: Advocate for ethical guidelines and standards in AI deployment, focusing on fairness, transparency, and accountability to mitigate potential biases and inequalities.
- Diverse Data and Interdisciplinary Research: Encourage the use of diverse datasets that represent varying demographics and economic sectors, and promote interdisciplinary research that combines insights from technology, sociology, and economics.
- Longitudinal Impact Studies: Invest in long-term studies to monitor the evolving impact of AI on employment, tracking changes in job roles, skills requirements, and worker adaptability.
- Public Awareness and Engagement: Launch public awareness campaigns to educate the broader population about AI's potential impacts and opportunities, facilitating a more informed and engaged workforce.
- Collaborative Policy Development: Work with policymakers, industry leaders, and educational institutions to develop and implement policies that support continuous learning, career mobility, and the transition of the workforce into AI-augmented roles.

Limitations and Future Directions

The analysis presented here is not without limitations. The complexity of AI's impact on employment cannot be captured fully through linear trends and aggregate statistics alone. Future research should:

- Integrate more nuanced data on AI's role within specific industries.
- Employ advanced statistical or machine learning models to uncover deeper correlations.
- Consider longitudinal studies to track the evolution of job roles in tandem with AI's advancement.

Concluding Remarks

As we stand on the cusp of a new era shaped by AI, the narrative that has emerged from our data is one of cautious optimism tempered by the recognition of impending challenges. The path forward is not one of resistance to change, but of adaptive strategies that embrace AI's potential while safeguarding the workforce against its disruptive forces. The enhanced recommendations provided aim to catalyze a proactive and inclusive approach, ensuring that the transition to an AI-driven labor market is equitable, ethical, and sustainable.

References

- OECD. (2022). OECD Employment Database. https://www.oecd.org/employment/emp/onlineoecdemploymentdatabase.htm
- Gupta, M. (n.d.). From Data Entry to CEO: The AI Job Threat Index. Kaggle. https://www.kaggle.com/datasets/manavgupta92/from-data-entry-to-ceo-the-ai-job-threat-index
- Frey, C. B., & Osborne, M. A. (2013).The Future of Employment: How susceptible are jobs to computerisation? Oxford Martin School, University of Oxford. https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf
- The Future of Jobs Report 2023. (2023, October 23). World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2023